Uvod u identifikabilnost

Aleksandar Hatzivelkos[1]

1. Uvod

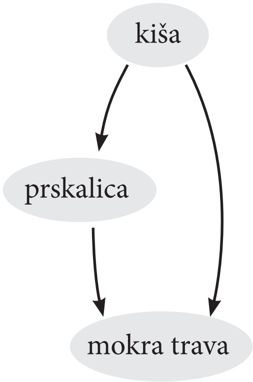
Teorija uzročnosti je teorija kojom se pomoću acikličkih usmjerenih grafova i pripadnih distribucija uvjetnih vjerojatnosti kroz koncept povećanja vjerojatnosti nekog događaja opisuju uzročno-posljedične veze. U takvom modelu, koncept identifikabilnosti predstavlja mogućnost izračuna uvjetnih vjerojatnosti iz neeksperimentalnih podataka. To znači da mjerljive veličine omogućavaju izračun vjerojatnosti isključivo iz podataka (distribucije vjerojatnosti) *a priori* vezanih uz uzročnu mrežu, a bez provođenja dodatnih eksperimenata. Radi se o svojstvu posebno važnom za uzročne modele u kojima nije jednostavno (ili je pak nemoguće) provesti eksperimente koji bi dali odgovore na tražena pitanja.

Uzročni modeli često se grade na osnovi nepotpunih statističkih podataka, pa tako ponekad u modelima postoje uzroci za koje znamo da postoje, znamo na koje varijable uzročno djeluju, no ne znamo u kojoj mjeri – tj. nemamo podatak o pripadnoj distribuciji uvjetne vjerojatnosti. Takve uzroke zovemo *nemjerljivima*. U ovom ćemo tekstu pokazati da su svi uzročni efekti identifikabilni u modelima u kojima ne postoje nemjerljivi uzroci, te ćemo vidjeti koji kriteriji osiguravaju identifikabilnost u modelima s nemjerljivim uzrocima.

2. Uzročna Bayesova mreža

Da bismo definirali pojam identifikabilnosti, prvo moramo opisati kontekst u kojemu se taj pojam koristi, tj. moramo definirati uzročnu Bayesovu mrežu. Uzročna Bayesova mreža sastoji se od skupa varijabli V i dviju matematičkih struktura definiranih nad njima. Prva je aciklički usmjereni graf G nad V. U grafu G svi su vrhovi varijable iz skupa V koje povezujemo usmjerenim strelicama. U grafu ne smiju postojati petlje, kao ni strelice koje vode iz nekog vrha u njega samoga. Druga struktura je distribucija uvjetnih vjerojatnosti definirana nad tvrdnjama o vrijednostima koje poprimaju varijable iz skupa V (uvjetno o vrijednostima varijabli iz kojih strelice vode u promatranu varijablu). Varijable u V mogu poprimati kontinuirane vrijednosti, no zbog jednostavnosti osnovna je teorija definirana nad varijablama koje poprimaju konačan broj vrijednosti.

[1] Aleksandar Hatzivelkos, Veleučilište Velika Gorica



Slika 1: Primjer jednostavnog direktnog acikličkog grafa

Ilustrirajmo opisanu strukturu primjerom. Na Slici 1. prikazan je usmjereni aciklički graf koji modelira odnose između tri binarne varijable, koje nam pak govore pada li kiša, je li prskalica uključena, te je li trava mokra. Modelom su opisane opće razumljive uzročnosti; ako pada kiša ili ako je uključena prskalica, tada kao posljedicu očekujemo da je trava mokra, te da, ukoliko pada kiša, nećemo uključivati prskalicu. No, da bi ovaj uzročni graf modelirao vjerojatnosne uzročne veze, moramo mu pridružiti i distribuciju vjerojatnosti, koja je dana u Tablici 1:

kiša	P(kiša)
0	0.75

prskalica	kiša	P(prskalica|kiša)
0	0	0.35
0	1	0.95

mokra trava	prskalica	kiša	P(mokra trava| prskalica,kiša)
0	0	0	0.99
0	0	1	0.19
0	1	0	0.05
0	1	1	0.02

Tablica 1: Distribucija uvjetnih vjerojatnosti

U opisivanju strukture grafa, uobičajeno koristimo *obiteljske* izraze. Tako za vrh V_i kažemo da je *roditelj* vrhu V_j ukoliko postoji strelica koja počinje u V_i a završava u V_j. Analogno, za vrh V_j kažemo da je *dijete* vrha V_i. Ukoliko, pak, od vrha V_i do vrha V_j postoji usmjereni put, tj niz strelica od kojih svaka počinje u vrhu u kojemu prethodna završava (osim prve strelice koja počinje u V_i), a posljednja završava u vrhu V_j, tada za vrh V_i kažemo da je *predak* vrha V_j, odnosno da je vrh V_j *potomak* vrha V_i. Jedino odstupanje od te razumljive terminologije predstavlja definicija po kojoj se svaka varijabla smatra svojim potomkom. Namjera je grafom nad skupom varijabli opisati uzročnu strukturu među varijablama. Strelica koja vodi iz vrha V_i u vrh V_j predstavlja neposredan uzročni utjecaj varijable V_i na varijablu V_j koji ne ovisi o drugim varijablama. Tada kažemo da je V_i *neposredan uzrok* od V_j.[2]

Distribucija vjerojatnosti P definirana je nad propozicijama oblika $X = x$, gdje je X varijabla iz skupa V, a x vrijednost u dometu varijable X. P je također definirana nad konjunkcijama, disjunkcijama i negacijama takvih propozicija. Odavde slijedi da će uvjetne vjerojatnosti nad takvim propozicijama biti dobro definirane kada uvjet ima strogo pozitivnu vjerojatnost. Uobičajena je upotreba pokrate u formi vjerojatnosnih tvrdnji koje sadrže samo varijable ili skupove varijabli, ali ne i njihove vrijednosti. Tako $P(X \mid Y) = P(X)$ koristimo kao pokratu za pisanje izraza

$$\forall x_1 \ldots \forall x_m \forall y_1 \ldots \forall y_n \left[P(X_1 = x_1 \wedge \ldots \wedge X_m = x_m \mid Y_1 = y_1 \wedge \ldots \wedge Y_n = y_n) = P(X_1 = x_1 \wedge \ldots \wedge X_m = x_m) \right].$$

2.1. Markovljev uvjet

Distribucija vjerojatnosti P nad V zadovoljava Markovljev uvjet ako je na sljedeći način povezana s grafom G:

$$\forall X \in V, \forall Y \subset V \setminus PO(X), P(X \mid RO(X) \wedge Y) = P(X \mid RO(X)), \qquad (1)$$

gdje je $PO(X)$ skup svih potomaka, a $RO(X)$ skup svih roditelja varijable X. Riječima, kažemo da zadovoljavanjem Markovljevog uvjeta roditelji svake varijable *zaklanjaju* svaki drugi podskup varijabli, osim potomaka varijable X.

Smisao uvjeta (1) je u tome da je poznavanje vrijednosti roditelja varijable X dovoljno kako bi se jednoznačno odredila vjerojatnost da će varijabla X poprimiti vrijednost x. Time zapravo tvrdimo da je varijabla X nezavisna od svih ostalih varijabli osim svojih potomaka, ukoliko joj poznajemo roditelje. Uzročni model, koji se sastoji od acikličkog usmjerenog grafa i distribucije vjerojatnosti koja zadovoljava Markovljev uvjet, nazivamo ***uzročnom Bayesovom mrežom***.

Neposredna posljedica zadovoljavanja Markovljevog uvjeta je mogućnost faktorizacije:

[2] S napomenom da je ta tvrdnja relativna s obzirom na skup varijabli V; uvođenjem neke nove varijable u model, kojom se detaljnije opisuje interakcija između varijabli V_i i V, V_j, može prestati biti neposredan uzrok.

Propozicija 2.1 *(o faktorizaciji)*

Neka distribucija vjerojatnosti *P* nad skupom varijabli *V* zadovoljava Markovljev uvjet, te neka je $V = \{X_1, X_2, ..., X_n\}$. Tada vrijedi:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i \mid RO(X_i)). \tag{2}$$

Dokaz. Bez smanjenja općenitosti, varijable u skupu *V* možemo indeksirati prema grafu *G* na način da indekse prvo dobiju varijable koje nisu ničiji potomci, te da za svake dvije varijable V_i, V_j, $i < j$, varijabla V_i može biti roditelj ili predak varijabli V_j, ali V_j ne može biti ni roditelj ni predak varijabli V_i.[3] Krenuvši od formule za uvjetnu vjerojatnost, $p(A,B) = p(A \mid B) \cdot p(B)$, iterativno raspisujemo:

$$\begin{aligned} P(X_1, X_2, ..., X_n) &= P(X_2, ..., X_n \mid X_1) \cdot P(X_1) \\ &= P(X_3, ..., X_n \mid X_1, X_2) \cdot P(X_2 \mid X_1) \cdot P(X_1) \\ &= P(X_4, ..., X_n \mid X_1, X_2, X_3) \cdot P(X_3 \mid X_1, X_2) \cdot P(X_2 \mid X_1) \cdot P(X_1) \\ &\cdots \\ &= P(X_n \mid X_1, ..., X_{n-1}) \cdot \cdots \cdot P(X_2 \mid X_1) \cdot P(X_1) \end{aligned}$$

Zadovoljavanje Markovljevog uvjeta sad osigurava da skup $RO(X_i)$ zaklanja sve ostale varijable o kojima uvjetno ovisi X_i, iz čega slijedi tvrdnja propozicije.

Promotrimo faktorizaciju koju, primjerice, generira graf *G* sa Slike 1. Kako u tom primjeru varijabla *K* (kiša) nema roditelja, varijabla *PR* (prskalica) za roditelja ima varijablu *K*, a varijabla *MT* (mokra trava) za roditelje ima varijable *K* i *PR*. Faktorizacija inducirana grafom, a iskazana jednadžbom (2), glasi

$$P(K, P, MT) = P(MT \mid K, PR) \cdot P(PR \mid K) \cdot P(K). \tag{3}$$

Želimo li sad, primjerice, izračunati vjerojatnost da je prskalica uključena, tj. $P(PR = 1)$, tada jednadžbu (3) trebamo prosumirati po svim vrijednostima koje poprimaju varijable *K* i *MT*, držeći vrijednost varijable *PR* fiksiranom na 1:

$$P(PR = 1) = \sum_{K,MT} P(MT \mid K, PR = 1) \cdot P(PR = 1 \mid K) \cdot P(K)$$

Izlučivanjem slijedi:

$$P(PR = 1) = \sum_{K} P(PR = 1 \mid K) \cdot P(K) \sum_{MT} P(MT \mid K, PR = 1)$$

Kako se unutarnja suma provodi po svim vrijednostima za *MT* uz fiksirane vrijednosti za *PR* i *K*, tako je ta suma jednaka 1. Slijedi:

$$P(PR = 1) = \sum_{K} P(PR = 1 \mid K) \cdot P(K),$$

[3] Poredak, dakle, ovisi o usmjerenju strelica u grafu i nije jedinstven.

što je i očekivano budući da prema grafu vrijednost vjerojatnosti za *PR* ovisi samo o roditelju *K*. Dakle:

$$
\begin{aligned}
P(PR=1) &= P(PR=1 \mid K=1)P(K=1) + P(PR=1 \mid K=0)P(K=0) \\
&= 0.05 \cdot 0.25 + 0.65 \cdot 0.75 \\
&= 0.5
\end{aligned}
$$

3. Identifikabilnost

Jedna od osnovnih koristi upotrebe uzročnih mreža jest mogućnost predviđanja, odnosno određivanja vrijednosti vjerojatnosti $P_X(Y)$, tj. vjerojatnosti da će varijabla Y poprimiti vrijednost y kada interveniramo na varijablu X fiksirajući je na vrijednost x. Tako označenu vjerojatnost $P_X(Y)$, koja nastaje *intervencijom* nad varijablom X, zovemo *uzročnim efektom*. Ona se u uzročnom grafu reprezentira brisanjem svih strelica koje vode u varijablu X koju pak fiksiramo na zadanu vrijednost. Posljedica u distribuciji uvjetnih vjerojatnosti jest da, nakon intervencije $X = x$, sve vjerojatnosti oblika $P(X = x \mid \cdots)$ poprimaju vrijednost 1, a sve vjerojatnosti oblika $P(X \neq x \mid \cdots)$ poprimaju vrijednost 0. Postavlja se pitanje možemo li iz neeksperimentalnih podataka *jednoznačno* odrediti vrijednost izraza $P_X(Y)$? U uzročnim mrežama u kojima su sve varijable mjerljive, odgovor na to pitanje je pozitivan – da, moguće je iz neeksperimentalnih podataka odrediti vrijednost ciljanih vjerojatnosti nakon intervencije. Ta je tvrdnja višekratno dokazana i neposredno slijedi iz činjenice da se u takvom grafu mogu računati sve (uvjetne) vjerojatnosti [3].

No, problem nastaje kada se matematički model susreće sa stvarnošću: naime, nije uvijek moguće dobiti podatke koji bi u potpunosti opisali uzročne ovisnosti između svih varijabli. Zapravo, u većini slučajeva poznat je tek dio uzročnih informacija. Za mnoge varijable nemamo dovoljno podataka da bismo uspostavili vjerojatnosne uzročne veze među njima, a ponekad nam nisu ni poznate sve varijable koje mogu utjecati na promatrani sustav. U toj situaciji uzročne grafove modeliramo uz pomoć *nemjerljivih* (*skrivenih, latentnih*) varijabli.

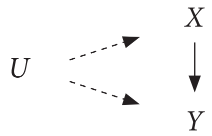
3.1 Neidentifikabilni uzročni efekt

Klasičan primjer takvog modela potječe iz 50-ih godina prošlog stoljeća kao posljedica rasprave o uzročnoj vezi između pušenja i obolijevanja od raka pluća. Visoka korelacija između ta dva događaja potaknula je postuliranje uzročne ovisnosti između pušenja (X) i obolijevanja od raka pluća (Y) koje bismo opisali jednostavnim uzročnim modelom (Slika 2).

$$X \longrightarrow Y$$

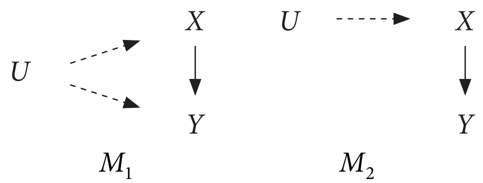
Slika 2: Jednostavan uzročni odnos pušenja i oboljenja od raka pluća

Iako ovu uzročnu vezu danas podrazumijevamo, uzročna interpretacija utvrđene korelacije na samom početku nije bila opće prihvaćena. Tako je, primjerice, britanski statističar R. A. Fisher ustvrdio da je predloženi model *„stara poznata greška kojom se korelacija tumači kao uzročna veza"*, te je ponudio model koji nudi alternativno objašnjenje korelacije između pušenja i obolijevanja od raka pluća [8].



Slika 3: Nemjerljivi zajednički uzrok pušenja i oboljenja od raka pluća

U tom je modelu korelacija između pušenja (*X*) i obolijevanja od raka pluća (*Y*) objašnjena pomoću genetske predispozicije koja povećava vjerojatnost oboljenja od raka pluća, ali i afinitet prema pušenju (U). S obzirom da se radi o 50-im godinama prošlog stoljeća, kada znanost još nije bila u stanju potvrditi ili opovrgnuti postojanje takve genetske predispozicije, varijabla U predstavljala je *nemjerljivi* zajednički uzrok varijabli *X* i *Y*. Taj je odnos u pripadnom grafu na Slici 3. istaknut isprekidanim strelicama, dakle uzročnom vezom koju ne možemo (vjerojatnosno) kvantificirati.



Slika 4: Mreže M_1 i M_2

Intuitivno, efekt varijable *X* na varijablu *Y*, tj. $P_X(Y)$, nije identifikabilan zato jer ne možemo sa sigurnošću reći je li obolijevanje od raka pluća rezultat pušenja ili pak nemjerljivog zajedničkog uzroka, genotipa koji simultano potiče i akciju i reakciju. Formalno, neidentifikabilnost ciljanog uzročnog efekta dokazujemo na sljedeći način: pokazat ćemo da je, za istu distribuciju mjerljivih varijabli, tj. uz poznati $P(X,Y)$, moguće konstruirati dvije *različite* uzročne Bayesove mreže, M_1 i M_2 (koje, u krajnosti, mogu biti i pravi podgrafovi mreže prikazane na Slici 3.) za koje je $P^{M_1}(X,Y) = P^{M_2}(X,Y)$.

Na Slici 4. prikazane su dvije takve mreže, M_1 i M_2, koje uzročne interpretacije dovode do ekstrema. Mrežom M_1 tako je prikazan model po kojemu obolijevanje od raka pluća uopće ne ovisi o pušenju, nego isključivo o genotipu osobe, dok je mrežom M_2 prikazan uzročni model po kojemu je obolijevanje od raka pluća isključiva posljedica pušenja, koje je pak posljedica genotipa (urođene sklonosti) osobe – dakle, model u kojemu genotip osobe ne uzrokuje neposredno rak pluća. Obje mreže moguće je konstruirati na način da je $P^{M_1}(X,Y) = P^{M_2}(X,Y)$, no s različitim distribucijama vjerojatnosti $P_X(Y)$. Napomenimo da za dokazivanje neidentifikabilnosti nije potrebno uzročnu interpretaciju dovesti do ekstrema – dovoljno je pokazati da na istome grafu postoje dvije različite distribucije vjerojatnosti koje daju tražene rezultate. Ipak, vrsta primjera koji interpretaciju uzročnosti dovode do ekstrema, efektnije ukazuje na suštinu neidentifikabilnosti kao koncepta.

X	U	$P^{M1}(X\mid U)$
0	0	0.3
0	1	0.7

Y	X	U	$P^{M1}(Y\mid X,U)$
0	0	0	0.75
0	0	1	0.25
0	1	0	0.75
0	1	1	0.25

Tablica 2: Distribucija vjerojatnosti kompatibilna s mrežom M_1

X	U	$P^{M2}(X\mid U)$
0	0	0.75
0	1	0.25

Y	X	U	$P^{M2}(Y\mid X,U)$
0	0	0	0.4
0	0	1	0.4
0	1	0	0.6
0	1	1	0.6

Tablica 3: Distribucija vjerojatnosti kompatibilna s mrežom M_2

Za početak, napomenimo da su sve veličine u ovim primjerima ilustrativnog karaktera, te ni na koji način ne opisuju stvarne ili pretpostavljene odnose između korištenih varijabli. Cilj je ovih primjera dokaz neidentifikabilnosti danog uzročnog efekta, a ne opisivanje stvarnih međuodnosa varijabli. U modelu ćemo pretpostaviti da je $P(U=0) = P(U=1) = 0.5$. Tablicom 2. dana je distribucija uvjetnih vjerojatnosti varijabli s grafa prikazanog na Slici 3., a koja je kompatibilna s grafom M_1 prikazanim na Slici 4. Uistinu, primijetimo da vrijednosti $P^{M_1}(Y \mid X,U)$ ne ovise o vrijednosti varijable *X*, što je kompatibilno s grafom u kojemu ne postoji strelica između varijabli *X* i *Y*.

S druge strane, Tablicom 3. dana je distribucija uvjetnih vjerojatnosti varijabli s grafa prikazanog na Slici 3., a koja je kompatibilna s grafom M_2 prikazanim na Slici 4. I ovdje možemo primijetiti da vrijednosti $P^{M_2}(Y \mid X,U)$ ne ovise o vrijednosti varijable *U*, što je kompatibilno s grafom u kojemu ne postoji strelica između varijabli *U* i *Y*.

Obje distribucije vjerojatnosti za mreže M_1 i M_2 računamo po formuli za faktorizaciju koja slijedi iz grafa na Slici 3:

$$P(X,Y,U) = P(Y \mid X,U) \cdot P(X \mid U) \cdot P(U) \tag{4}$$

Da bismo izračunali traženu vjerojatnost $P(X,Y)$, jednadžbu (4) prosumiramo po U. Tako za mrežu M_1 imamo:

$$\begin{aligned} P^{M_1}(X=0,Y=0) &= \sum_U P^{M_1}(Y=0 \mid X=0,U) P^{M_1}(X=0 \mid U) P^{M_1}(U) \\ &= 0.75 \cdot 0.3 \cdot 0.5 + 0.25 \cdot 0.7 \cdot 0.5 = 0.2 \end{aligned}$$

Na isti način, sumiranjem jednadžbe (4) po *U* za mrežu M_2, dobivamo:

$$\begin{aligned} P^{M_2}(X=0,Y=0) &= \sum_U P^{M_2}(Y=0 \mid X=0,U) P^{M_2}(X=0 \mid U) P^{M_2}(U) \\ &= 0.4 \cdot 0.75 \cdot 0.5 + 0.4 \cdot 0.25 \cdot 0.5 = 0.2 \end{aligned}$$

Analogno, i za ostale vrijednosti varijabli *X* i *Y* u obje mreže dobivamo iste vrijednosti, što konačno rezultira istom distribucijom:

X	Y	$P(X,Y)$
0	0	0.2
0	1	0.3
1	0	0.3
1	1	0.2

Tablica 4: Distribucija vjerojatnosti $P(X,Y)$ u M_1 i M_2

Distribuciju vjerojatnosti varijable *Y* nakon intervencije na varijablu *X*, $P_X(Y)$, također računamo sumiranjem jednadžbe (4) po *U*. No, pri tome je vrijednost varijable *X* intervencijom fiksirana na zadanu vrijednost *x*, pa je svaka vjerojatnost oblika $P(X = x \mid \cdots)$ jednaka 1. Slijedi:

$$\begin{aligned} P_X(Y) &= \sum_U P(Y \mid X=x,U) P(X=x \mid U) P(U) \\ &= \sum_U P(Y \mid X=x,U) P(U) \end{aligned}$$

Promotrimo npr. intervenciju $X=1$ i pripadnu distribuciju vjerojatnosti za *Y* u obje mreže:

$$P_{X=1}^{M_1}(Y=0) = \sum_U P^{M_1}(Y \mid X=1,U)P^{M_1}(U)$$
$$= 0.75 \cdot 0.5 + 0.25 \cdot 0.5 = 0.5$$

$$P_{X=1}^{M_2}(Y=0) = \sum_U P^{M_2}(Y \mid X=1,U)P^{M_2}(U)$$
$$= 0.6 \cdot 0.5 + 0.6 \cdot 0.5 = 0.6$$

Ovime je dokazano da uzročni efekt $P_X(Y)$ u mreži danoj na Slici 3. nije identifikabilan.

3.2 Kriteriji identifikabilnosti

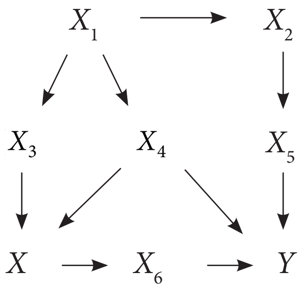
Nakon što smo pokazali na koji način uzročni efekt u uzročnoj mreži može biti neidentifikabilan, vrijeme je da ponudimo odgovor i na pitanje utvrđivanja identifikabilnosti. Prvi sustavno obrađeni kriteriji su „back-door" i „front-door" kriteriji kojima se osigurava identifikabilnost uzročnog efekta $P_X(Y)$.

Definicija 3.1

Za skup varijabli Z kažemo da zadovoljava kriterij stražnjih vrata (back-door kriterij) u odnosu na uređeni par varijabli (X,Y) u usmjerenom acikličkom grafu Z ako vrijedi:

(I) *niti jedna varijabla iz Z nije potomak od X*

(II) *Z blokira svaki put između X i Y u kojemu strelica vodi u X.*



Slika 5: Primjer za kriterij stražnjih vrata: skup $\{X_3, X_4\}$ zadovoljava kriterij za par varijabli (X,Y).

Ime je kriterij dobio po drugom uvjetu definicije, po kojemu se Z vizualno u grafu nalazi „straga" u odnosu na X. Primjerice, u grafu prikazanom na Slici 5. skupovi $Z_1 = \{X_3, X_4\}$ i $Z_2 = \{X_4, X_5\}$ zadovoljavaju kriterij za par varijabli (X,Y), dok ga skup $Z_3 = \{X_4\}$ ne zadovoljava budući da varijabla X_4 ne blokira put $(X, X_3, X_1, X_2, X_5, Y)$.

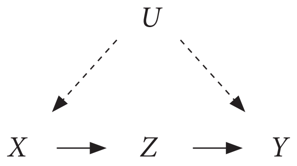
Teorem 3.1 *(o kriteriju stražnjih vrata)*

Ako skup Z zadovoljava kriterij stražnjih vrata u odnosu na uređeni par varijabli (X,Y), tada je uzročni efekt varijable X na varijablu Y identifikabilan, i vrijedi

$$P_X(Y) = \sum_Z P(Y \mid X,Z) \cdot P(Z). \tag{5}$$

Dokaz teorema provodi se analizom formule za računanje vrijednosti $P_X(Y)$. Bez detaljnijeg ulaženja u formalni raspis intervencije (koja je u ovom tekstu opisana na konceptualnoj, ali ne i na strogo formalnoj razini), dokaz se temelji na definiciji kriterija stražnjih vrata – skupom Z pokrivene su sve varijable koje utječu na varijablu X pa se izlučivanjem sumacija u formuli (2) svodi na sumiranje po vrijednostima varijabli iz skupa Z, odakle slijedi tvrdnja.

Promotrimo sada primjer prikazan na Slici 5., no dodatno pretpostavimo da su varijable $\{X_1, X_2, X_3, X_4, X_5\}$ nemjerljive, te ih označimo novom zajdničkom nemjerljivom varijablom U. Varijabli X_6 također promijenimo oznaku, te je označimo sa Z. Tako dobiveni graf prikazan je na sl. 6.



Slika 6: Primjer za kriterij prednjih vrata: uzročni efekt $P_X(Y)$ je identifikabilan uz poznavanje Z.

Distribucija vjerojatnosti dana je formulom (2) i glasi:

$$P(X,Y,Z,U) = P(Y \mid Z,U) \cdot P(Z \mid X) \cdot P(X \mid U) \cdot P(U). \tag{6}$$

Intervencijom na vrijednost varijable X faktor $P(X \mid U)$ poprima vrijednost 1, pa jednadžba prelazi u

$$P_X(Y,Z,U) = P(Y \mid Z,U) \cdot P(Z \mid X) \cdot P(U).$$

Sumiranje po vrijednostima varijabli Z i U daje:

$$P_X(Y) = \sum_Z P(Z \mid X) \sum_U P(Y \mid Z,U) \cdot P(U). \tag{7}$$

Cilj nam je eliminirati varijablu U iz desne strane jednadžbe (7). Za tu ćemo svrhu upotrijebiti dvije uzročne nezavisnosti koje slijede iz grafa sa Slike 6:

$$P(U \mid Z,X) = P(U \mid X), \tag{8}$$

$$P(Y \mid X,Z,U) = P(Y \mid Z,U). \tag{9}$$

Odavde nam slijedi:

$$\begin{aligned}\sum\nolimits_U P(Y\mid Z,U)\cdot P(U) &= \sum\nolimits_{X'}\sum\nolimits_U P(Y\mid Z,U)P(U\mid X')P(X')\\ &= \sum\nolimits_{X'}\sum\nolimits_U P(Y\mid X',Z,U)P(U\mid X',Z)P(X')\\ &= \sum\nolimits_{X'} P(Y\mid X',Z)P(X'),\end{aligned}$$

čime smo jednadžbu (7) iskazali isključivo pomoću mjerljivih varijabli:

$$P_X(Y)=\sum_Z P(Z\mid X)\sum_{X'} P(Y\mid X',Z)\cdot P(X'). \tag{10}$$

Svi faktori s desne strane jednadžbe (10) izračunljivi su iz mjerljivih veličina, pa je onda to i $P_X(Y)$. Dakle, traženi uzročni efekt je identifikabilan, čak i uz odstustvo kriterija stražnjih vrata, ukoliko imamo međuvarijablu *Z* koja zadovoljava uvjete iskazane jednadžbama (8) i (9).

Jednadžba (10) može se interpretirati kao upotreba kriterija stražnjih vrata u dva koraka. U prvom koraku odredimo uzročni efekt varijable *X* na varijablu *Z* – kako nema drugih puteva kroz stražnja vrata do varijable *Z* osim onog iz *X*, imamo:

$$P_X(Z)=P(Z\mid X).$$

Nakon toga računamo uzročni efekt *Z* na *Y*, koji nije samo $P(Y\mid Z)$, budući da imamo put kroz stražnja vrata od varijable *Z* do *Y*: $Z\leftarrow X\leftarrow U\rightarrow Y$. No, kako na tom putu leži *X* koji generira uvjetne neovisnosti[4] (8) i (9), tako on može odigrati ulogu posrednika u kriteriju stražnjih vrata koji nam omogućava izračun $P_Z(Y)=\sum\nolimits_{X'} P(Y\mid X',Z)\cdot P(X')$. Kombiniranjem ovih uzročnih efekata slijedi formula (10). Time smo opisali *kriterij prednjih vrata*:

Definicija 3.2

Za skup varijabli Z kažemo da zadovoljava kriterij prednjih vrata u odnosu na uređeni par varijabli (X,Y) *u usmjerenom acikličkom grafu G ako:*

(I) Z *leži na svim usmjerenim putovima od X do Y*

(II) *ne postoji put kroz stražnja vrata od X do Z*

(III) *X blokira sve putove kroz stražnja vrata od Z do Y.*

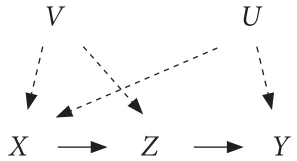
[4] To se svojstvo naziva *d-separacijom*, i daje grafički kriterij za određivanje uvjetnih neovisnosti u danom uzročnom grafu. Kako u okviru ovog teksta nismo mogli pokriti i to područje teorije uzročnosti, predlažemo daljnje čitanje [4].

Teorem 3.2 *(o kriteriju prednjih vrata)*

Ako skup Z zadovoljava kriterij prednjih vrata u odnosu na uređeni par varijabli (X,Y) i ako je $P(Z \mid X) > 0$, tada je uzročni efekt varijable X na varijablu Y identifikabilan, i vrijedi

$$P_X(Y) = \sum_Z P(Z \mid X) \sum_{X'} P(Y \mid X', Z) \cdot P(X').$$

Nakon što smo se upoznali s kriterijima za određivanje identifikabilnosti uzročnog efekta, vratimo se klasičnom primjeru pušenja i obolijevanja od raka pluća, prikazanom grafom na Slici 2. Kao što smo već pokazali, uzročni efekt varijable X (pušenja) na varijablu Y (rak pluća) u tom grafu je neidentifikabilan. Rješenje za taj problem može se naći u proširivanju uzročnog modela na onaj prikazan Slikom 6. Naravno, uz drugačiju interpretaciju danih varijabli. Takvom mrežom modeliramo situaciju u kojoj imamo nemjerljivi zajednički uzrok, tj. genetsku predispoziciju (varijabla U) koju uzrokuje pušenje (varijabla X) i rak pluća (varijabla Y), te varijablu Z koja mjeri naslagu katrana u plućima. U tom modelu pušenje uzrokuje rak pluća isključivo kroz katran u plućima. Ukoliko se uzročna veza između pušenja i raka pluća modelira na taj način, uzročni efekt je identifikabilan kroz zadovoljavanje kriterija prednjih vrata.



Slika 7: Primjer neidentifikabilnog uzročnog efekta $P_X(Y)$ *u mreži s dva nemjerljiva uzroka.*

Naravno, ni taj model nije ostao bez odgovora. Ukoliko se pak tom modelu uvede još jedan nemjerljivi zajednički uzrok, ovaj put varijabli X i Z, kao što je prikazano na Slici 7., varijable X i Y više ne zadovoljavaju kriterij prednjih vrata – više nije zadovoljen uvjet *(ii)* iz definicije, tj. kroz varijablu V postoji put kroz stražnja vrata od X do Z. Kao i u primjeru uzročne mreže sa Slike 3, moguće je konstruirati distribuciju vjerojatnosti kojom se dokazuje neidentifikabilnost uzročnog efekta $P_X(Y)$. Naravno da je sada već teže dati uvjerljivo objašnjenje o (drugoj) genetskoj predispoziciji koja uzrokuje sklonost ka pušenju i (neovisno od pušenja) nakupljanje katrana u plućima. Isto tako, danas je genom čovjeka poznat do te mjere da teze o nemjerljivom zajedničkom genetskom uzroku pušenja i raka pluća nisu znanstveno vjerodostojne. No, sama povijest rješavanja ovog problema ukazuje na važnost koncepta identifikabilnosti u uzročnim mrežama.

Literatura

1. A. Hatzivelkos *Što je uzročnost*, Poučak 50, (2012.)
2. C. Hitchcock *Probabilistic Causation*, The Stanford Encyclopedia of Philosophy (Winter 2011. Edition)
3. J. Pearl *Graphical models, causality, and intervention*, Statist. Sci. Volume 8, Number 3, 266-269. (1993.)
4. J. Pearl *Causality: models, reasoning and inference*, Cambridge University Press, (2000.)
5. J. Pearl *Direct and Indirect Effects*, Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, San Francisco, CA: Morgan Kaufmann, 411-420, (2001.)
6. J. Pearl, J. Tian *On the Identifcation of Causal Effects*, UCLA Cognitive Systems Laboratory, Technical Report (R-290-L), (2003.)
7. C. R. Shalizi *Advanced Data Analysis from an Elementary Point of View*, (2011.) http://www.cscs.umich.edu/.17ex : crshalizi/weblog/cat_36-402.html
8. M. Valtorta *Identifiability in Causal Bayesian Networks: A Gentle Introduction*, Cybernetics and Systems: An International Journal, Volume 39, Issue 4, (2008.)